ALPHA (1,2) FUCOSYLTRANSFERASE SYNGENES FOR USE IN THE PRODUCTION OF FUCOSYLATED OLIGOSACCHARIDES

FIELD OF THE INVENTION

The invention provides compositions and methods for producing purified oligosaccharides, in particular certain fucosylated oligosaccharides that are typically found in human milk.

BACKGROUND OF THE INVENTION

Human milk contains a diverse and abundant set of neutral and acidic oligosaccharides. More than 130 different complex oligosaccharides have been identified in human milk, and their structural diversity and abundance is unique to humans. Although these molecules may not be utilized directly by infants for nutrition, they nevertheless serve critical roles in the establishment of a healthy gut microbiome, in the prevention of disease, and in immune function. Prior to the invention described herein, the ability to produce human milk oligosaccharides (HMOS) inexpensively was problematic. For example, their production through chemical synthesis was limited by stereo-specificity issues, precursor availability, product impurities, and high overall cost. As such, there is a pressing need for new strategies to inexpensively manufacture large quantities of HMOS.

SUMMARY

The disclosure features an efficient and economical method for producing fucosylated oligosaccharides. Such production of a fucosylated oligosaccharide is accomplished using an isolated nucleic acid comprising a sequence encoding a lactose-utilizing α (1,2) fucosyltransferase gene product (e.g., polypeptide or protein), which is operably linked to one or more heterologous control sequences that direct the production of the recombinant fucosyltransferase gene product in a host production bacterium such as *Escherichia coli* (*E. coli*).

In an aspect, there is provided a method for producing a fucosylated oligosaccharide in a bacterium comprising providing bacterium comprising an exogenous lactose-utilizing α(1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme comprises at least 75% identity to SEQ ID NO: 1 or at least 75% identity to SEQ ID NO:3, and

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;

(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;

(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;

(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or

(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19.

In another aspect, there is provided a method for producing a fucosylated oligosaccharide in a bacterium comprising providing bacterium comprising an exogenous lactose-utilizing α(1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme comprises

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;

(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;

(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;

(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or

(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19.

The present disclosure provides novel α (1,2) fucosyltransferases (also referred to herein as α(1,2) FTs) that utilize lactose and catalyzes the transfer of an L-fucose sugar from a GDP-fucose donor substrate to an acceptor substrate in an alpha-1,2-linkage. In a preferred embodiment, the acceptor substrate is an oligosaccharide. The α(1,2) fucosyltransferases identified and described herein are useful for expressing in host bacterium for the production of human milk oligosaccharides (HMOS), such as fucosylated oligosaccharides. Exemplary

fucosylated oligosaccharides produced by the methods described herein include 2'-fucosyllactose (2'FL), lactodifucotetraose (LDFT), lacto-N-fucopentaose I (LNF I), or lacto-N-difucohexaose I (LDFH I). The "α(1,2) fucosyltransferases" disclosed herein encompasses the amino acid sequences of the α(1,2) fucosyltransferases and the nucleic acid sequences that encode the α(1,2) fucosyltransferases, as well as variants and fragments thereof that exhibit α(1,2) fucosyltransferase activity. Also within the disclosure is a nucleic acid construct comprising an isolated nucleic acid encoding a lactose-accepting α (1,2) fucosyltransferase enzyme, said nucleic acid being optionally operably linked to one or more heterologous control sequences that direct the production of the enzyme in a host bacteria production strain.

In another aspect, there is provided a nucleic acid construct comprising an isolated nucleic acid encoding a lactose-utilizing α(1,2) fucosyltransferase enzyme for the production of said enzyme in a bacterium production strain, wherein the amino acid sequence of said enzyme comprises at least 75% identity to SEQ ID NO: 1 or at least 75% identity to SEQ ID No:3, and

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;

(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;

(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;

(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or

(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19; and

wherein said bacterium further comprises reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated adenosine-5'-triphosphate (ATP)-dependent intracellular protease, an inactivated endogenous *lacA* gene, or any combination thereof.

In another aspect, there is provided a nucleic acid construct comprising an isolated nucleic acid encoding a lactose-utilizing α(1,2) fucosyltransferase enzyme for the production of said enzyme in a bacteriumproduction strain, wherein the amino acid sequence of said enzyme comprises

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;

(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;

(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;

(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or

(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19; and

wherein said bacterium further comprises reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated adenosine-5'-triphosphate (ATP)-dependent intracellular protease, an inactivated endogenous *lacA* gene, or any combination thereof. In another aspect, there is provided an isolated bacterium comprising an isolated nucleic acid encoding a lactose-accepting α (1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme encoded by said nucleic acid comprises at least 75% identity to SEQ ID NO: 1 or at least 75% identity to SEQ ID No: 3, and

(i) at least 75% identity to SEQ ID NO: 11;

(ii) at least 75% identity to SEQ ID NO: 12;

(iii) at least 75% identity to SEQ ID NO: 13;

(iv) at least 75% identity to SEQ ID NO: 14; or

(v) at least 75% identity to SEQ ID NO: 19; and

wherein the isolated bacterium further comprising a α(1,3) fucosyltransferase enzyme and/or a α(1,4) fucosyltransferase enzyme.

In another aspect, there is provided an isolated bacterium comprising an isolated nucleic acid encoding a lactose-accepting α (1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme encoded by said nucleic acid comprises

(i) at least 75% identity to SEQ ID NO: 11;

(ii) at least 75% identity to SEQ ID NO: 12;

(iii) at least 75% identity to SEQ ID NO: 13;

(iv) at least 75% identity to SEQ ID NO: 14; or

(v) at least 75% identity to SEQ ID NO: 19; and

wherein the isolated bacterium further comprising a α(1,3) fucosyltransferase enzyme and/or a α(1,4) fucosyltransferase enzyme.

The amino acid sequence of the lactose-accepting α(1,2) fucosyltransferases described herein is at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identity to *Helicobacter pylori* 26695 alpha-(1,2) fucosyltransferase (futC or SEQ ID NO: 1). Preferably, the lactose-accepting α(1,2) fucosyltransferases described herein is at least 22% identical to *H. pylori* FutC, or SEQ ID NO: 1.

The amino acid sequence of the lactose-accepting α(1,2) fucosyltransferases described herein is at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identity to *Bacteroides vulgatus* alpha-(1,2) fucosyltransferase (FutN or SEQ ID NO: 3). Preferably, the lactose-accepting α(1,2) fucosyltransferases described herein is at least 25% identical to *B. vlugatos* FutN, or SEQ ID NO: 3.

Alternatively, the exogenous α (1,2) fucosyltransferase preferably comprises at least at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identity to any one of the novel α (1,2) fucosyltransferases disclosed herein, for example, to the amino acid sequences in Table 1.

Exemplary α(1,2) fucosyltransferases include, but are not limited to, *Prevotella melaninogenica* FutO, *Clostridium bolteae* FutP, *Clostridium bolteae* +13 FutP, *Lachnospiraceae sp.* FutQ, *Methanosphaerula palustris* FutR, *Tannerella sp.* FutS, *Bacteroides caccae* FutU, *Butyrivibrio* FutV, *Prevotella sp.* FutW, *Parabacteroides johnsonii* FutX, *Akkermansia muciniphilia* FutY, *Salmonella enterica* FutZ, *Bacteroides sp.* FutZA.

For example, the α(1,2) fucosyltransferases comprise the amino acid sequences comprising any one of the following: *Prevotella melaninogenica* FutO (SEQ ID NO: 10), *Clostridium bolteae* FutP (SEQ ID NO: 11), *Clostridium bolteae* +13 FutP (SEQ ID NO: 292), *Lachnospiraceae sp.* FutQ (SEQ ID NO: 12), *Methanosphaerula palustris* FutR (SEQ ID NO: 13), *Tannerella sp.* FutS (SEQ ID NO: 14), *Bacteroides caccae* FutU (SEQ ID NO: 15), *Butyrivibrio* FutV (SEQ ID NO: 16), *Prevotella sp.* FutW (SEQ ID NO: 17), *Parabacteroides johnsonii* FutX (SEQ ID NO: 18), *Akkermansia muciniphilia* FutY (SEQ ID NO: 19), *Salmonella enterica* FutZ (SEQ ID NO: 20), and *Bacteroides sp.* FutZA (SEQ ID NO: 21), or a functional variant or fragment thereof. Other exemplary α(1,2) fucosyltransferases include any of the enzymes listed in Table 1, or functional variants or fragments thereof.

The present disclosure features a method for producing a fucosylated oligosaccharide in a bacterium by providing bacterium that express at least one exogenous lactose-utilizing α(1,2) fucosyltransferase. The amino acid sequence of the exogenous lactose-utilizing α(1,2) fucosyltransferase is preferably at least 22% identical to *H. pylori* FutC or at least 25% identical to *B. vulgatus* FutN. The bacterium also expresses one or more exogenous lactose-utilizing α(1,3) fucosyltransferase enzymes and/or one or more exogenous lactose-utilizing α(1,4) fucosyltransferase enzymes. The combination of fucosyltransferases expressed in the production bacterium is dependent upon the desired fucosylated oligosaccharide product. The method disclosed herein further includes retrieving the fucosylated oligosaccharide from said bacterium or from a culture supernatant of said bacterium.

Examples of suitable α(1,3) fucosyltransferase enzymes include, but are not limited to *Helicobacter pylori* 26695 *futA* gene (GenBank Accession Number HV532291 (GI:365791177), incorporated herein by reference), *H. hepaticus* Hh0072, *H.pylori* 11639 FucT, and *H.pylori* UA948 FucTa (e.g., GenBank Accession Number AF194963 (GI:28436396), incorporated herein by reference)(Rasko, D. A., Wang, G., Palcic, M. M. & Taylor, D. E. J Biol Chem 275, 4988-4994 (2000)). Examples of suitable a(1,4) fucosyltransferase enzymes include, but are not limited to *H.pylori* UA948 FucTa (which has has relaxed acceptor specificity and is able to generate both α(1,3)- and α(1,4)-fucosyl linkages). An example of an enzyme possessing only α(1,4) fucosyltransferase activity is given by the FucT III enzyme from *Helicobacter pylori* strain DMS6709 (e.g., GenBank Accession Number AY450598.1 (GI:40646733), incorporated herein by reference) (S. Rabbani, V. Miksa, B. Wipf, B. Ernst, *Glycobiology* **15**, 1076-83 (2005).)

The disclosure also features a nucleic acid construct or a vector comprising a nucleic acid enconding at least one α (1,2) fucosyltransferase or variant, or fragment thereof, as described herein. The vector can further include one or more regulatory elements, e.g., a heterologous promoter. By "heterologous" is meant that the control sequence and protein-encoding sequence originate from different bacterial strains. The regulatory elements can be operably linked to a gene encoding a protein, a gene construct encoding a fusion protein gene, or a series of genes linked in an operon in order to express the fusion protein. The disclosure comprises an isolated recombinant cell, e.g., a bacterial cell containing an aforementioned nucleic acid molecule or vector. The nucleic acid is optionally integrated into the genome of the host bacterium. In some embodiments, the nucleic acid construct also further comprises one or more α(1,3) fucosyltransferases and/or α(1,4) fucosyltransferases. Alternatively, the α (1,2) fucosyltransferase also exhibits α(1,3) fucosyltransferase and/or α(1,4) fucosyltransferase activity.

The bacterium utilized in the production methods described herein is genetically engineered to increase the efficiency and yield of fucosylated oligosaccharide products. For example, the host production bacterium is characterized as having a reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated ATP-dependent intracellular protease, an inactivated *lacA*, or a combination thereof. In one embodiment, the bacterium is characterized as having a reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated ATP-dependent intracellular protease, and an inactivated *lacA*.

As used herein, an "inactivated" or "inactivation of a" gene, encoded gene product (*i.e.*, polypeptide), or pathway refers to reducing or eliminating the expression (*i.e.*, transcription or translation), protein level (*i.e.*, translation, rate of degradation), or enzymatic activity of the gene, gene product, or pathway. In the instance where a pathway is inactivated, preferably one enzyme or polypeptide in the pathway exhibits reduced or negligible activity. For example, the enzyme in the pathway is altered, deleted or mutated such that the product of the pathway is produced at low levels compared to a wild-type bacterium or an intact pathway. Alternatively, the product of the pathway is not produced. Inactivation of a gene is achieved by deletion or mutation of the gene or regulatory elements of the gene such that the gene is no longer transcribed or translated. Inactivation of a polypeptide can be achieved by deletion or mutation of the gene that encodes the gene product or mutation of the polypeptide to disrupt its activity. Inactivating mutations include

to export excess lactose from its cytoplasm (Danchin A. Cells need safety valves. Bioessays 2009, Jul;31(7):769-73.), thus greatly facilitating purposeful manipulations of the *E.coli* intracellular lactose pool.

The disclosure also provides methods for increasing intracellular levels of GDP-fucose in a bacterium by manipulating the organism's endogenous colanic acid biosynthesis pathway. This increase is achieved through a number of genetic modifications of endogenous *E. coli* genes involved either directly in colanic acid precursor biosynthesis, or in overall control of the colanic acid synthetic regulon. Particularly preferred is inactivation of the genes or encoded polypeptides that act in the colanic acid synthesis pathway after the production of GDP-fucose (the donor substrate) and before the generation of colanic acid. Exemplary colanic acid synthesis genes include, but are not limited to: a *wcaJ* gene, (e.g., GenBank Accession Number (amino acid) BAA15900 (GI:1736749), incorporated herein by reference), a *wcaA* gene (e.g., GenBank Accession Number (amino acid) BAA15912.1 (GI:1736762), incorporated herein by reference), a *wcaC* gene (e.g., GenBank Accession Number (amino acid) BAE76574.1 (GI:85675203), incorporated herein by reference), a *wcaE* gene (e.g., GenBank Accession Number (amino acid) BAE76572.1 (GI:85675201), incorporated herein by reference), a *wcaI* gene (e.g., GenBank Accession Number (amino acid) BAA15906.1 (GI:1736756), incorporated herein by reference), a *wcaL* gene (e.g., GenBank Accession Number (amino acid) BAA15898.1 (GI:1736747), incorporated herein by reference), a *wcaB* gene (e.g., GenBank Accession Number (amino acid) BAA15911.1 (GI:1736761), incorporated herein by reference), a *wcaF* gene (e.g., GenBank Accession Number (amino acid) BAA15910.1 (GI:1736760), incorporated herein by reference), a *wzxE* gene (e.g., GenBank Accession Number (amino acid) BAE77506.1 (GI:85676256), incorporated herein by reference), a *wzxC* gene, (e.g., GenBank Accession Number (amino acid) BAA15899 (GI:1736748), incorporated herein by reference), a *wcaD* gene, (e.g., GenBank Accession Number (amino acid) BAE76573 (GI:85675202), incorporated herein by reference), a *wza* gene (e.g., GenBank Accession Number (amino acid) BAE76576 (GI:85675205), incorporated herein by reference), a *wzb* gene (e.g., GenBank Accession Number (amino acid) BAE76575 (GI:85675204), incorporated herein by reference), and a *wzc* gene (e.g., GenBank Accession Number (amino acid) BAA15913 (GI:1736763), incorporated herein by reference).

Preferably, a host bacterium, such as *E. coli*, is genetically engineered to produce a human milk oligosaccharide by the inactivation of the *wcaJ* gene, which encoding the UDP-glucose lipid carrier transferase. The inactivation of the *wcaJ* gene can be by deletion of the

characterized by the deletion of the endogenous β-galactosidase gene, an exogenous β-galactosidase gene is introduced to the bacterium. For example, a plasmid expressing an exogenous β-galactosidase gene is introduced to the bacterium, or recombined or integrated into the host genome. For example, the exogenous β-galactosidase gene is inserted into a gene that is inactivated in the host bacterium, such as the *lon* gene.

The exogenous b-galactosidase gene is a functional b-galactosidase gene characterized by a reduced or low leve of b-galactosidase activity compared to β-galactosidase activity in wild-type bacteria lacking any genetic manipulation. Exemplary β-galactosidase genes include *E. coli lacZ* and β-galactosidase genes from any of a number of other organisms (e.g., the *lac4* gene of *Kluyveromyces lactis* (e.g., GenBank Accession Number M84410 (GI:173304), incorporated herein by reference) that catalyzes the hydrolysis of b-galactosides into monosaccharides. The level of β-galactosidase activity in wild-type *E. coli* bacteria is, for example, 6,000 units. Thus, the reduced β-galactosidase activity level encompassed by engineered host bacterium of the present disclosure includes less than 6,000 units, less than 5,000 units, less than 4,000 units, less than 3,000 units, less than 2,000 units, less than 1,000 units, less than 900 units, less than 800 units, less than 700 units, less than 600 units, less than 500 units, less than 400 units, less than 300 units, less than 200 units, less than 100 units, or less than 50 units. Low, functional levels of β-galactosidase include β-galactosidase activity levels of between 0.05 and 1,000 units, e.g., between 0.05 and 750 units, between 0.05 and 500 units, between 0.05 and 400 units, between 0.05 and 300 units, between 0.05 and 200 units, between 0.05 and 100 units, between 0.05 and 50 units, between 0.05 and 10 units, between 0.05 and 5 units, between 0.05 and 4 units, between 0.05 and 3 units, or between 0.05 and 2 units of β-galactosidase activity. For unit definition and assays for determining β-galactosidase activity, see Miller JH, Laboratory CSH. Experiments in molecular genetics. Cold Spring Harbor Laboratory Cold Spring Harbor, NY; 1972; (incorporated herein by reference). This low level of cytoplasmic β-galactosidase activity is not high enough to significantly diminish the intracellular lactose pool. The low level of β-galactosidase activity is very useful for the facile removal of undesired residual lactose at the end of fermentations.

Optionally, the bacterium has an inactivated *thyA* gene. Preferably, a mutation in a *thyA* gene in the host bacterium allows for the maintenance of plasmids that carry *thyA* as a selectable marker gene. Exemplary alternative selectable markers include antibiotic resistance genes such as BLA (beta-lactamase), or *proBA* genes (to complement a *proAB*

host strain proline auxotropy) or *purA* (to complement a *purA* host strain adenine auxotrophy).

The *E. coli* bacterium comprises the genotype Δ*ampC*::P_{trp}^{B}*cI*, Δ(*lacI-lacZ*)::*FRT*, P_{lacIq}*lacY*$^{+}$, Δ*wcaJ*::FRT, *thyA*::Tn*10*, Δ*lon*:(*npt3*, *lacZ*$^{+}$), Δ*lacA*, and also comprises any one of the exogenous α(1,2) fucosyltransferases described herein.

The bacterium comprising these characteristics is cultured in the presence of lactose. In some cases, the method further comprises culturing the bacterium in the presence of tryptophan and in the absence of thymidine. The fucosylated oligosaccharide is retrieved from the bacterium (*i.e.*, a cell lysate) or from a culture supernatant of the bacterium.

The disclosure provides a purified fucosylated oligosaccharide produced by the methods described herein. The fucosylated oligosaccharide is purified for use in therapeutic or nutritional products, or the bacterium is used directly in such products. The fucosylated oligosaccharide produced by the engineered bacterium is 2'-fucosyllactose (2'-FL) or lactodifucotetraose (LDFT). The new alpha 1,2-fucosyltransferases are also useful to synthesize HMOS of larger molecular weight bearing alpha 1,2 fucose moieties, e.g., lacto-N-fucopentaose (LNF I) and lacto-N-difucohexaose (LDFH I). For example, to produce LDFT, the host bacterium is engineered to express an exogenous α (1,2) fucosyltransferase that also possesses α (1,3) fucosyltransferase activity, or an exogenous α (1,2) fucosyltransferase and an exogenous α (1,3) fucosyltransferase. For the production of LNF I and LDFH I, the host bacterium is engineered to express an exogenous α (1,2) fucosyltransferase that also possesses α (1,3) fucosyltransferase activity and/or α (1,4) fucosyltransferase activity, or an exogenous α (1,2) fucosyltransferase, an exogenous α (1,3) fucosyltransferas, and an exogenous α (1,4) fucosyltransferase.

A purified fucosylated oligosaccharide produced by the methods described above is also within the disclosure. The purified oligosaccharide (2'-FL) obtained at the end of the process is a white/slightly off-white, crystalline, sweet powder. For example, an engineered bacterium, bacterial culture supernatant, or bacterial cell lysate according to the disclosure comprises 2'-FL, LDFT, LNF I or LDFH I produced by the methods described herein, and does not substantially comprise a other fucosylated oligosaccharides prior to purification of the fucosylated oligosaccharide products from the cell, culture supernatant, or lysate. As a general matter, the fucosylated oligosaccharide produced by the methods contains a negligible amount of 3-FL in a 2'-FL-containing cell, cell lysate or culture, or supernatant, e.g., less than 1% of the level of 2'-FL or 0.5% of the level of 2'-FL. Moreover, the

fucosylated oligosaccharide produced by the methods described herein also have a minimal amount of contaminating lactose, which can often be co-purified with the fucosylated oligosaccharide product, such as 2'FL. This reduction in contaminating lactose results from the reduced level of β-galactosidase activity present in the engineered host bacterium.

A purified oligosaccharide, e.g., 2'-FL, LDFT, LNF I, or LDFH I, is one that is at least 90%, 95%, 98%, 99%, or 100% (w/w) of the desired oligosaccharide by weight. Purity is assessed by any known method, e.g., thin layer chromatography or other chromatographic techniques known in the art. The disclosure includes a method of purifying a fucosylated oligosaccharide produced by the genetically engineered bacterium described above, which method comprises separating the desired fucosylated oligosaccharide (e.g., 2'-FL) from contaminants in a bacterial cell lysate or bacterial cell culture supernatant of the bacterium.

The oligosaccharides are purified and used in a number of products for consumption by humans as well as animals, such as companion animals (dogs, cats) as well as livestock (bovine, equine, ovine, caprine, or porcine animals, as well as poultry). For example, a pharmaceutical composition comprises purified 2'-FL and a pharmaceutically-acceptable excipient that is suitable for oral administration. Large quantities of 2'-FL are produced in bacterial hosts, e.g., an *E. coli* bacterium comprising an exogenous α (1,2) fucosyltransferase gene.

A method of producing a pharmaceutical composition comprising a purified human milk oligosaccharide (HMOS) is carried out by culturing the bacterium described above, purifying the HMOS produced by the bacterium, and combining the HMOS with an excipient or carrier to yield a dietary supplement for oral administration. These compositions are useful in methods of preventing or treating enteric and/or respiratory diseases in infants and adults. Accordingly, the compositions are administered to a subject suffering from or at risk of developing such a disease.

The disclosure also provides methods of identifying an α (1,2) fucosyltransferase gene capable of synthesizing fucosylated oligosaccharides in a host bacterium, *i.e.*, 2'-fucosyllactose (2'-FL) in *E. coli*. The method of identifying novel lactose-utilizing, α(1,2)fucosyltransferase enzyme comprises the following steps:

1) performing a computational search of sequence databases to define a broad group of simple sequence homologs of any known, lactose-utilizing α(1,2)fucosyltransferase;

2) using the list from step (1), deriving a search profile containing common sequence and/or structural motifs shared by the members of the list;

3) searching sequence databases, using a derived search profile based on the common sequence or structural motif from step (2) as query, and identifying a candidate sequences, wherein a sequence homology to a reference lactose-utilizing α(1,2)fucosyltransferase is a predetermined percentage threshold;

4) compiling a list of candidate organisms, said organisms being characterized as expressing α(1,2)fucosyl- glycans in a naturally-occurring state;

5) selecting candidate sequences that are derived from candidate organisms to generate a list of candidate lactose-utilizing enzymes;

6) expressing the candidate lactose- utilizing enzyme in a host organism; and

7) testing for lactose- utilizing α(1,2)fucosyltransferase activity, wherein detection of the desired fucosylated oligosaccharide product in said organism indicates that the candidate sequence comprises a novel lactose- utilizing α(1,2)fucosyltransferase. In another embodiment, the search profile is generated from a multiple sequence alignment of the amino acid sequences of more than one enzyme with known α(1,2)fucosyltransferase activity. The database search can then be designed to refine and iteratively search for novel α(1,2)fucosyltransferases with significant sequence similarlity to the multiple sequence alignment query.

The disclosure provides a method of treating, preventing, or reducing the risk of infection in a subject comprising administering to said subject a composition comprising a purified recombinant human milk oligosaccharide, wherein the HMOS binds to a pathogen and wherein the subject is infected with or at risk of infection with the pathogen. The infection is caused by a Norwalk-like virus or *Campylobacter jejuni.* The subject is preferably a mammal in need of such treatment. The mammal is, e.g., any mammal, e.g., a human, a primate, a mouse, a rat, a dog, a cat, a cow, a horse, or a pig. In a preferred embodiment, the mammal is a human. For example, the compositions are formulated into animal feed (e.g., pellets, kibble, mash) or animal food supplements for companion animals, e.g., dogs or cats, as well as livestock or animals grown for food consumption, e.g., cattle, sheep, pigs, chickens, and goats. Preferably, the purified HMOS is formulated into a powder (e.g., infant formula powder or adult nutritional supplement powder, each of which is mixed with a liquid such as water or juice prior to consumption) or in the form of tablets, capsules or pastes or is incorporated as a component in dairy products such as milk, cream, cheese, yogurt or kefir, or as a component in any beverage, or combined in a preparation containing live microbial cultures intended to serve as probiotics, or in prebiotic preparations to enhance the growth of beneficial microorganisms either *in vitro* or *in vivo*.

Polynucleotides, polypeptides, and oligosaccharides of the disclosure are purified and/or isolated. Purified defines a degree of sterility that is safe for administration to a human subject, e.g., lacking infectious or toxic agents. Specifically, as used herein, an "isolated" or "purified" nucleic acid molecule, polynucleotide, polypeptide, protein or oligosaccharide, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. For example, purified HMOS compositions are at least 60% by weight (dry weight) the compound of interest. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight the compound of interest. Purity is measured by any appropriate standard method, for example, by column chromatography, thin layer chromatography, or high-performance liquid chromatography (HPLC) analysis. For example, a "purified protein" refers to a protein that has been separated from other proteins, lipids, and nucleic acids with which it is naturally associated. Preferably, the protein constitutes at least 10, 20, 50, 70, 80, 90, 95, 99-100% by dry weight of the purified preparation.

Similarly, by "substantially pure" is meant an oligosaccharide that has been separated from the components that naturally accompany it. Typically, the oligosaccharide is substantially pure when it is at least 60%, 70%, 80%, 90%, 95%, or even 99%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated.

By "isolated nucleic acid" is meant a nucleic acid that is free of the genes which, in the naturally-occurring genome of the organism from which the DNA of the disclosure is derived, flank the gene. The term covers, for example: (a) a DNA which is part of a naturally occurring genomic DNA molecule, but is not flanked by both of the nucleic acid sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner, such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Isolated nucleic acid molecules according to the present disclosure further include molecules produced synthetically, as well as any nucleic acids that have been altered chemically and/or that have modified backbones.

A "heterologous promoter" is a promoter which is different from the promoter to which a gene or nucleic acid sequence is operably linked in nature.

The term "overexpress" or "overexpression" refers to a situation in which more factor is expressed by a genetically-altered cell than would be, under the same conditions, by a wild type cell. Similarly, if an unaltered cell does not express a factor that it is genetically altered to produce, the term "express" (as distinguished from "overexpress") is used indicating the wild type cell did not express the factor at all prior to genetic manipulation.

The terms "treating" and "treatment" as used herein refer to the administration of an agent or formulation to a clinically symptomatic individual afflicted with an adverse condition, disorder, or disease, so as to effect a reduction in severity and/or frequency of symptoms, eliminate the symptoms and/or their underlying cause, and/or facilitate improvement or remediation of damage. The terms "preventing" and "prevention" refer to the administration of an agent or composition to a clinically asymptomatic individual who is susceptible to a particular adverse condition, disorder, or disease, and thus relates to the prevention of the occurrence of symptoms and/or their underlying cause.

By the terms "effective amount" and "therapeutically effective amount" of a formulation or formulation component is meant a nontoxic but sufficient amount of the formulation or component to provide the desired effect.

The transitional term "comprising," which is synonymous with "including," "containing," or "characterized by," is inclusive or open-ended and does not exclude additional, unrecited elements or method steps. By contrast, the transitional phrase "consisting of" excludes any element, step, or ingredient not specified in the claim. The transitional phrase "consisting essentially of" limits the scope of a claim to the specified materials or steps "and those that do not materially affect the basic and novel characteristic(s)" of the claimed disclosure.

The host organism used to express the lactose-accepting fucosyltransferase gene is typically the enterobacterium *Escherichia coli* K12 (*E. coli*). *E. coli* K-12 is not considered a human or animal pathogen nor is it toxicogenic. E. coli K-12 is a standard production strain of bacteria and is noted for its safety due to its poor ability to colonize the colon and establish infections (see, e.g., epa.gov/oppt/biotech/pubs/fra/fra004.htm). However, a variety of bacterial species may be used in the oligosaccharide biosynthesis methods, e.g., *Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum*, *or Xanthomonas campestris.* Bacteria of the genus *Bacillus* may also be used, including *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus thermophilus*,

Bacillus laterosporus, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus*, *and Bacillus circulans*. Similarly, bacteria of the genera *Lactobacillus* and *Lactococcus* may be modified using the methods of this disclosure, including but not limited to *Lactobacillus acidophilus, Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii*, *and Lactococcus lactis. Streptococcus thermophiles and Proprionibacterium freudenreichii* are also suitable bacterial species for the disclosure described herein. Also included as part of this disclosure are strains, modified as described here, from the genera *Enterococcus* (e.g., *Enterococcus faecium* and *Enterococcus thermophiles*), *Bifidobacterium* (e.g., *Bifidobacterium longum*, *Bifidobacterium infantis*, and *Bifidobacterium bifidum*), *Sporolactobacillus* spp., *Micromomospora* spp., *Micrococcus* spp., *Rhodococcus* spp., and *Pseudomonas* (e.g., *Pseudomonas fluorescens* and *Pseudomonas aeruginosa*). Bacteria comprising the characteristics described herein are cultured in the presence of lactose, and a fucosylated oligosaccharide is retrieved, either from the bacterium itself or from a culture supernatant of the bacterium. The fucosylated oligosaccharide is purified for use in therapeutic or nutritional products, or the bacteria are used directly in such products. A suitable production host bacterial strain is one that is not the same bacterial strain as the source bacterial strain from which the fucosyltransferase-encoding nucleic acid sequence was identified.

Other features and advantages of the disclosure will be apparent from the following description of the preferred embodiments thereof, and from the claims. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. All published foreign patents and patent applications cited herein are incorporated herein by reference. Genbank and NCBI submissions indicated by accession number cited herein are incorporated herein by reference. All other published references, documents, manuscripts and scientific literature cited herein are incorporated herein by reference. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

What is claimed is:

1. A method for producing a fucosylated oligosaccharide in a bacterium comprising providing bacterium comprising an exogenous lactose-utilizing α(1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme comprises at least 75% identity to SEQ ID NO: 1 or at least 75% identity to SEQ ID NO:3, and

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;
(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;
(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;
(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or
(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19.

2. A method for producing a fucosylated oligosaccharide in a bacterium comprising providing bacterium comprising an exogenous lactose-utilizing α(1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme comprises

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;
(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;
(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;
(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or
(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19.

3. The method of claim 1 or 2, further comprising retrieving the fucosylated oligosaccharide from said bacterium or from a culture supernatant of said bacterium.

4. The method of any one of the preceding claims, wherein said fucosylated oligosaccharide comprises 2'-fucosyllactose (2'-FL), lactodifucotetraose (LDFT), Lacto-N-fucopentaose I (LNF I), or lacto-N-difucohexaose I (LDFH I).

5. The method of any one of the preceding claims, wherein the bacterium further comprises an exogenous lactose-utilizing α(1,3) fucosyltransferase enzyme and/or an exogenous lactose-utilizing α(1,4) fucosyltransferase enzyme.

6. The method of claim 5, wherein the exogenous lactose-utilizing α(1,3) fucosyltransferase enzyme comprises (i) a *Helicobacter pylori* 26695 *futA* gene; or (ii) a Helicobacter pylori UA948 FucTa gene or a Helicobacter pylori strain DMS6709 FucT III gene.

7. The method of any one of the preceding claims, wherein said bacterium further comprises a reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated ATP-dependent intracellular protease, an inactivated *lacA*, or a combination thereof.

8. The method of claim 7, wherein

(i) said method further comprises culturing said bacterium in the presence of tryptophan and in the absence of thymidine;

(ii) said reduced level of β-galactosidase activity comprises a deleted or inactivated endogenous lacZ gene and/or a deleted or inactivated endogenous lacl gene of said bacterium;

(iii) said reduced level of β-galactosidase activity comprises a deleted or inactivated endogenous lacZ gene and/or a deleted or inactivated endogenous lacI gene of said bacterium, wherein said reduced level of β-galactosidase activity further comprises an exogenous lacZ gene or variant thereof, wherein said exogenous lacZ gene or variant thereof comprises an β-galactosidase activity level less than wild-type bacterium;

(iv) said reduced level of β-galactosidase activity comprises an activity level less than wild-type bacterium;

(v) said reduced level of β-galactosidase activity comprises an activity level less than wild-type bacterium, wherein said reduced level of β-galactosidase activity comprises less than 6,000 units of β-galactosidase activity;

(vi) said reduced level of β-galactosidase activity comprises an activity level less than wild-type bacterium, wherein said reduced level of β-galactosidase activity comprises less than 1,000 units of β-galactosidase activity;

(vii) said bacterium comprises a lacIq gene promoter immediately upstream of a lacY gene;

(viii) said defective colonic acid synthesis pathway comprises the inactivation of the wcaJ gene of said bacterium;

(ix) said inactivated ATP-dependent intracellular protease is a null mutation, inactivating mutation, or deletion of an endogenous Ion gene;

(x) said inactivated ATP-dependent intracellular protease is an inactivating mutation of an endogenous lon gene, wherein said inactivating mutation of an endogenous lon gene comprises the insertion of a functional E. coli $lacZ^+$ gene;

(xi) said bacterium further comprises a functional lactose permease gene;

(xii) said bacterium further comprises a functional lactose permease gene, wherein said bacterium comprises E. coli lacY;

(xiii) said bacterium further comprises an exogenous E. coli rcsA or E. coli rcsB gene;

(xiv) said bacterium further comprises a mutation in the thyA gene;

(xv) said bacterium accumulates intracellular lactose in the presence of exogenous lactose;

(xvi) said bacterium accumulates intracellular GDP-fucose; or

(xvii) said bacterium comprises the genotype ΔampC::$P_{trp}{}^{B}$cI, Δ(lacI-lacZ)::FRT,P_{lacIq}lacY^+, ΔwcaJ::FRT, thyA::Tn10, Δlon:(npt3, $lacZ^{+)}$.

9. The method of any one of the preceding claims, wherein said bacterium is a member of the *Bacillus, Pantoea, Lactobacillus, Lactococcus, Streptococcus, Proprionibacterium, Enterococcus, Bifidobacterium, Sporolactobacillus*, *Micromomospora*, *Micrococcus, Rhodococcus*, or *Pseudomonas* genus, preferably wherein said production strain is selected from the group consisting of *Bacillus licheniformis, Bacillus subtilis*, *Bacillus coagulans*, *Bacillus thermophilus*, *Bacillus laterosporus*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus*, *and Bacillus circulans, Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum*, *Xanthomonas campestris Lactobacillus acidophilus, Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii*, *Lactococcus lactis, Streptococcus thermophiles, Proprionibacterium freudenreichii, Enterococcus faecium, Enterococcus thermophiles*), *Bifidobacterium longum*, *Bifidobacterium infantis*, *Bifidobacterium bifidum, Pseudomonas fluorescens* and *Pseudomonas aeruginosa.*

10. The method of any of the preceding claims, wherein said bacterium comprises a nucleic acid construct comprising an isolated nucleic acid encoding said α(1,2) fucosyltransferase enzyme.

11. A nucleic acid construct comprising an isolated nucleic acid encoding a lactose-utilizing α(1,2) fucosyltransferase enzyme for the production of said enzyme in a bacterium production strain, wherein the amino acid sequence of said enzyme comprises at least 75% identity to SEQ ID NO: 1 or at least 75% identity to SEQ ID No:3, and

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;
(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;
(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;
(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or
(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19; and

wherein said bacterium further comprises reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated adenosine-5'-triphosphate (ATP)-dependent intracellular protease, an inactivated endogenous *lacA* gene, or any combination thereof.

12. A nucleic acid construct comprising an isolated nucleic acid encoding a lactose-utilizing α(1,2) fucosyltransferase enzyme for the production of said enzyme in a bacteriumproduction strain, wherein the amino acid sequence of said enzyme comprises

(i) an amino acid sequence having at least 75% identity to SEQ ID NO: 11;
(ii) an amino acid sequence having at least 75% identity to SEQ ID NO: 12;
(iii) an amino acid sequence having at least 75% identity to SEQ ID NO: 13;
(iv) an amino acid sequence having at least 75% identity to SEQ ID NO: 14; or
(v) an amino acid sequence having at least 75% identity to SEQ ID NO: 19; and

wherein said bacterium further comprises reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated adenosine-5'-triphosphate (ATP)-dependent intracellular protease, an inactivated endogenous *lacA* gene, or any combination thereof.

13. The method or construct of claims 10 or 11-12, wherein said nucleic acid is operably linked to one or more heterologous control sequences that direct the production of the

enzyme in the bacterium, preferably wherein said heterologous control sequence comprises a bacterial promoter and operator, a bacterial ribosome binding site, a bacterial transcriptional terminator, a plasmid selectable marker, and/or an origin of replication.

14. An isolated bacterium comprising an isolated nucleic acid encoding a lactose-accepting α (1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme encoded by said nucleic acid comprises at least 75% identity to SEQ ID NO: 1 or at least 75% identity to SEQ ID No: 3, and

(i) at least 75% identity to SEQ ID NO: 11;

(ii) at least 75% identity to SEQ ID NO: 12;

(iii) at least 75% identity to SEQ ID NO: 13;

(iv) at least 75% identity to SEQ ID NO: 14; or

(v) at least 75% identity to SEQ ID NO: 19; and

wherein the isolated bacterium further comprising a α(1,3) fucosyltransferase enzyme and/or a α(1,4) fucosyltransferase enzyme.

15. An isolated bacterium comprising an isolated nucleic acid encoding a lactose-accepting α (1,2) fucosyltransferase enzyme, wherein the amino acid sequence of said enzyme encoded by said nucleic acid comprises

(i) at least 75% identity to SEQ ID NO: 11;

(ii) at least 75% identity to SEQ ID NO: 12;

(iii) at least 75% identity to SEQ ID NO: 13;

(iv) at least 75% identity to SEQ ID NO: 14; or

(v) at least 75% identity to SEQ ID NO: 19; and

wherein the isolated bacterium further comprising a α(1,3) fucosyltransferase enzyme and/or a α(1,4) fucosyltransferase enzyme.

16. The method, construct, or isolated bacterium of any one of claims 1 to 10 or 11 to 15, wherein said α(1,2) fucosyltransferase enzyme comprises any one selected from *Tannerella sp.* FutS, *Clostridium bolteae* +13 FutP, *Methanosphaerula palustries* FutR, *Akkermansia muciniphilia* FutY, *Clostridium bolteae* FutP, or a functional variant or fragment thereof.

17. The method, construct, or isolated bacterium of any one of claims 1 to 10 or 11 to 16, wherein said bacterium is *Escherichia coli*.

18. The method or isolated bacterium of any one of claims 1 to 10 or 13 to 17, wherein said bacterium further comprises reduced level of β-galactosidase activity, a defective colanic acid synthesis pathway, an inactivated adenosine-5'-triphosphate (ATP)-dependent intracellular protease, an inactivated endogenous *lacA* gene, or any combination thereof.

19. The construct of any one of claims 11 to 13 or 18, wherein said bacterium further comprises the genotype Δ*ampC*::P_{trp}^{B}*cI*, Δ(*lacI-lacZ*)::*FRT*, P_{lacIq}*lacY*$^{+}$, Δ*wcaJ*::FRT, *thyA*::Tn*10*, Δ*lon*:(*npt3*, *lacZ*$^{+}$).

20. The method, construct, or isolated bacterium of any one of claims 1 to 10, 11 to 19, wherein said enzyme is not a mutant enzyme.

21. The method, construct, or isolated bacterium of claim 20, wherein said enzyme naturally occurs in a bacterium that naturally exists in the gastrointestinal tract.

22. The method, construct, or isolated bacterium of any one of claims 1 to 19, wherein the enzyme comprises

(i) the amino acid sequence of SEQ ID NO: 11;
(ii) the amino acid sequence of SEQ ID NO: 12;
(iii) the amino acid sequence of SEQ ID NO: 13;
(iv) the amino acid sequence of SEQ ID NO: 14; or
(v) the amino acid sequence of SEQ ID NO: 19.

23. The method, construct, or isolated bacterium of any one of claims 1 to 19, wherein the enzyme comprises an amino acid sequence having at least 75% identity to SEQ ID NO: 13.

24. The method, construct, or isolated bacterium of claim 23, wherein the enzyme comprises the amino acid sequence of SEQ ID NO: 13.

25. The method, construct, or isolated bacterium of any one of claims 1-2 or 11-12 or 14-15, wherein the amino acid sequence of said enzyme comprises at least 95% identity to SEQ ID NO: 1 or at least 95% identity to SEQ ID NO:3, and

(i) an amino acid sequence having at least 95% identity to SEQ ID NO: 11;

(ii) an amino acid sequence having at least 95% identity to SEQ ID NO: 12;

(iii) an amino acid sequence having at least 95% identity to SEQ ID NO: 13;

(iv) an amino acid sequence having at least 95% identity to SEQ ID NO: 14; or

(v) an amino acid sequence having at least 95% identity to SEQ ID NO: 19.

26. The method, construct, or isolated bacterium of any one of claims 1-2 or 11-12 or 14-15, wherein the amino acid sequence of said enzyme comprises

(i) an amino acid sequence having at least 95% identity to SEQ ID NO: 11;

(ii) an amino acid sequence having at least 95% identity to SEQ ID NO: 12;

(iii) an amino acid sequence having at least 95% identity to SEQ ID NO: 13;

(iv) an amino acid sequence having at least 95% identity to SEQ ID NO: 14; or

(v) an amino acid sequence having at least 95% identity to SEQ ID NO: 19.